



# Genomic Biology Resources

National Center for Biotechnology Information ■ National Library of Medicine ■ National Institutes of Health Department of

**NCBI** genomic biology starts at Entrez Genomes. Entrez Genomes provides access to a collection of genomic sequences representing over 900 organisms, including sequences from bacteria, archaea, eukaryota, viruses, viroids and organelles. Both completely sequenced genomes and those for which sequencing is in progress are included. Combined with related resources at NCBI, Entrez Genomes provides scientists with a powerful gateway to the analysis and comparison of complete genomes. Supplementing the interactive database and tools are FTP sites for the bulk download of sequence data and tables of annotations.

Entrez Genomes [www.ncbi.nlm.nih.gov/genome/guide/human/](http://www.ncbi.nlm.nih.gov/genome/guide/human/)

Genomes FTP <ftp://ncbi.nlm.nih.gov/genomes/>

## Entrez Genomes Resource Pages

Genome Resource Pages, described below, serve as collection points for genome data resources for individual organisms or groups of organisms with links to genomic sequence data, precomputed analyses such as the protein alignments shown in BLINK reports, and tools such as BLAST. Other resources available for each organism or genome are listed to the right of the resource page description using two-letter codes (refer to the key in the "Tools and Cross-Species Resources" section on the back).

**Human Genome Resources Guide** [www.ncbi.nlm.nih.gov/genome/guide/human/](http://www.ncbi.nlm.nih.gov/genome/guide/human/)

BL MV TP LL RS TA HM UG PE HG

Contains bulletins and progress reports concerning the Human Genome Project and provides centralized access to tools ranging from the human genome Map Viewer and Human Genome BLAST, to the clone registry.

**Mouse Genome Resources Guide** [www.ncbi.nlm.nih.gov/genome/guide/mouse/](http://www.ncbi.nlm.nih.gov/genome/guide/mouse/)

BL MV TP LL RS TA HM UG PE HG

Collects a variety of services running the gamut from information on mouse strains and annotation projects to genomic display and analysis tools.

**Rat Genome Resources Guide** [www.ncbi.nlm.nih.gov/genome/guide/R\\_norvegicus.html](http://www.ncbi.nlm.nih.gov/genome/guide/R_norvegicus.html)

BL MV TP LL RS TA UG PE HG

Brings together information on diverse rat-related resources including sequence, mapping, clone information, and pointers to strain and mutation resources.

**Fruit Fly Genome Resources Guide** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/7227.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/7227.html)

BL MV TP LL RS UG PE HG

Provides the assembled and annotated genome sequence of the euchromatic arms of the five *Drosophila melanogaster* chromosomes via the Map Viewer. The sequence was determined in a collaboration between Celera and the Berkeley *Drosophila* Genome Project.

**Mosquito Genome Resources Guide** [www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map\\_search?chr=agambiae.inf](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/map_search?chr=agambiae.inf)

BL MV TP TA UG PE

Allows access to the genome assembly of the *A. gambiae*, with links to taxonomy information.

**Plant Genomes Central** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/PlantList.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/PlantList.html)

BL MV TP RS UG PE HG

Provides access to large-scale plant genomic and EST sequencing data, with links to corresponding taxonomic information. For the 'Large-scale sequencing projects' and the 'Genetic maps' groups, the organism name links to Map Viewer overviews. For the 'Large-scale EST sequencing project' group, the organism name links to the taxonomy page at NCBI.

**Zebrafish Genome Resources** [www.ncbi.nlm.nih.gov/genome/guide/D\\_rerio.html](http://www.ncbi.nlm.nih.gov/genome/guide/D_rerio.html)

BL MV TP LL RS TA UG PE HG

Links to a wide spectrum of sites related to the zebrafish genome including genetic maps and mutagenic projects.

**Malaria Genetics and Genomics** [www.ncbi.nlm.nih.gov/projects/Malaria/](http://www.ncbi.nlm.nih.gov/projects/Malaria/)

BL MV TP RS

Provides resources including organism specific BLAST databases (*Plasmodium falciparum* only, all *Plasmodium*), genome maps, linkage markers, and information about genetic studies.

**Retroviruses** [www.ncbi.nlm.nih.gov/retroviruses](http://www.ncbi.nlm.nih.gov/retroviruses)

BL LL RS

A collection of retrovirus-oriented resources, including an HIV-1 sequence annotation tool, graphical representations of retroviral genomes, and a Genotyping tool, which uses the BLAST algorithm to determine the genotype of a query sequence.

**Microbial Genomes** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html)

BL CG TP RS TA

Shows graphical representations of complete bacterial genomes that can be viewed in their entirety or explored in progressively greater detail. A "ProtTable" of protein coding genes is provided for each genome. There are also links to a "TaxTable," showing the distribution of BLAST protein homologs by taxa (sequences grouped by superkingdom), and to a distribution of BLAST protein homologs by 3-D structure (sequences with known structure).

**Organelle Genome Resources** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/organelles.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/organelles.html)

BL RS

Provides a list of complete mitochondrial and plastid genomes, and presents tools that can be used to analyze these sequences.

**Virus Reference Genomes** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/viruses.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/viruses.html)

BL RS

Includes a general introduction to viruses and a variety of genome retrieval tools, such as the Viral Genomes Finder. Virus-specific analysis programs, such as the Clusters of Related Viral Proteins (CRP), are also available.

**Eukaryote Genomes Projects** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/EG\\_T.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/EG_T.html)

Displays a list of eukaryote sequencing efforts for which genomic data is publicly available.

# Tools and Cross-Species Resources

## **BL Genomic BLAST** [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)

Allows sequence queries against completed and/or unfinished genomes of specific organisms or groups of organisms.

## **MV Entrez Map Viewer** [www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewHelp.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/MapViewHelp.html)

View and search an organism's complete genome, display chromosome maps, and zoom to progressively greater levels of detail, down to the sequence data for a region of interest. The maps available vary by organism, and are described in the Map Viewer "data and search tips". Multiple maps can be displayed and are aligned to one another based on shared markers.

## **MM Model Maker**

Allows the construction of transcript models using novel combinations of putative exons derived from *ab initio* predictions or from the alignment of GenBank transcripts, including ESTs and NCBI RefSeqs, to the NCBI human genome assembly.

## **EV Evidence Viewer**

Displays the alignments to a genomic contig of RefSeq transcripts, GenBank mRNAs, known or potential transcripts, and ESTs supporting a gene model. The EV produces a graphical summary of the alignments that indicates the coordinate range of the gene model on the genomic contig and the areas of alignment of the transcripts, including as ESTs, on separate tracks. Following the graphical summary are detailed exon-by-exon alignments of all of the transcript sequences against the genomic contig, including flanking genomic sequence for each exon to show the presence of absence of splice sites.

## **CG Clusters of Orthologous Groups (COGs)** [www.ncbi.nlm.nih.gov/COG/](http://www.ncbi.nlm.nih.gov/COG/)

A database of proteins encoded in complete genomes. COGs currently consists of over 3300 clusters from 43 microbial genomes including representatives from the bacteria, archaea and one eukaryote, *Saccharomyces cerevisiae*. In addition, proteins from two higher eukaryotic genomes, *C. elegans* and *D. melanogaster*, have been assigned to COGs. COGs are integrated with Entrez Genomes where proteins are linked to their respective COGs, and COG data is included in several display options.

## **TP TaxPlot** [www.ncbi.nlm.nih.gov/cgi-bin/Entrez/taxik2](http://www.ncbi.nlm.nih.gov/cgi-bin/Entrez/taxik2)

A tool for 3-way comparisons of genomes on the basis of the sequences of the proteins they encode. Pre-computed BLAST results are used to plot a point for each predicted protein in a reference genome, based on the best alignment with proteins in each of the two genomes being compared.

## **LL LocusLink** [www.ncbi.nlm.nih.gov/LocusLink/index.html](http://www.ncbi.nlm.nih.gov/LocusLink/index.html)

A single query interface to curated sequence and descriptive information about genetic loci. It presents information on official nomenclature, aliases, sequence accessions, phenotypes, EC numbers, MIM numbers, UniGene clusters, gene homology, map locations, and related web sites.

## **RS RefSeq** [www.ncbi.nlm.nih.gov/LocusLink/refseq.html](http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html)

The NCBI Reference Sequence project (RefSeq) provides reference sequence standards for chromosomes, mRNAs, other RNAs, and Proteins. RefSeq standards provide a foundation for the functional annotation of the human and other genomes and serve as a stable reference for gene characterization, mutation analysis, expression studies, and polymorphism discovery.

## **TA Trace Archive** [www.ncbi.nlm.nih.gov/Traces/trace.cgi?](http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?) [www.ncbi.nlm.nih.gov/blast/tracemb.html](http://www.ncbi.nlm.nih.gov/blast/tracemb.html)

A repository of raw sequence traces generated by large-scale sequencing projects. In the case of projects that rely on a Whole Genome Shotgun (WGS) strategy, the Trace Archive will be the sole source of raw sequence data. Trace BLAST uses MegaBLAST for queries against the trace data. Discontiguous MegaBLAST allows rapid and sensitive cross-species searches.

## **HM Human-Mouse Homology Map** [www.ncbi.nlm.nih.gov/Homology/](http://www.ncbi.nlm.nih.gov/Homology/)

Displays synteny maps for human and mouse. Comparisons of the human genome vs. the MGD Genetic map and the Whitehead/MRC RH map are available.

## **UG UniGene** [www.ncbi.nlm.nih.gov/UniGene/](http://www.ncbi.nlm.nih.gov/UniGene/)

Non-redundant gene-based clusters of GenBank mRNA sequences and ESTs for more than a dozen organisms including human, cow, mouse, rat, frog, zebrafish, fruitfly, mosquito, barley, rice, thale cress, maize, and wheat. The UniGene databases are updated weekly with new EST sequences, and bimonthly with newly characterized sequences.

## **PE ProtEST** [www.ncbi.nlm.nih.gov/UniGene/protest.shtml](http://www.ncbi.nlm.nih.gov/UniGene/protest.shtml)

Pre-computed BLAST alignments between protein sequences from 8 model organisms, including *H. sapiens*, *M. musculus*, *R. norvegicus*, *D. melanogaster*, *C. elegans*, *S. cerevisiae*, *A. thaliana*, and *E. coli*, and the 6-frame translations of UniGene nucleotide sequences. ProtEST links are displayed in UniGene reports in the section on model organism protein similarities.

## **HG HomoloGene** [www.ncbi.nlm.nih.gov/HomoloGene](http://www.ncbi.nlm.nih.gov/HomoloGene)

Curated and calculated homologs for genes as represented by UniGene, LocusLink, and more recently, by genomic sequences, and assembled contigs using whole genome shotgun (WGS) reads. The calculated homologs are the result of nucleotide sequence comparisons between each pair of organisms, using EST and mRNA sequences from UniGene, as well as genomic sequences and contigs.

